International Journal of Engineering & Technology, 8 (1.7) (2019) 162-170

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

ORSUM: a Machine Learning Approach for Intelligent Transportation

Imran Medi[1]*, Aida Mustapha[2], Vinothini Kasinathan[3]

[1]Faculty of Computing, Engineering and Technology, Asia Pacific University of Innovation and Technology, Malaysia
[2]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia,
86400 Batu Pahat, Johor, Malaysia
**Corresponding author E-mail: imran.medi@apu.edu.my*

Abstract

Ride-hailing applications such as Uber and Lyft responds to requests similar to taxis calls, whereby a driver drives from any location nearby to fetch passengers to make profit. This paper presents ORSUM, a ride-sharing application which allows drivers and commuters to share the cost of a journey should be able to provide the desired convenience and costs for moving about in a city. In this study, the proposed application capitalizes on machine learning approach to learn users' daily travel patterns and recommend "ride buddies" for which the ride is to be shared with. ORSUM has four distinct modules; Orsum Machine Learning module (ORSUMML), Azure Database, Orsum Web Application, and Orsum Application. The machine learning module runs as a standalone Django web applica-tion that is separate from the Orsum Web Application and only interacts with the Orsum Web Application. ORSUMML is developed using the Python based Sci-kit learn or C# based Accord.Net. Evaluation of ORSUM showed high user acceptance rate to the application, comparable to existing applications such as Uber and Lyft.

Keywords: *Ride-Sharing; Carpooling; K-Means Clustering; Machine Learning*

1. Introduction

Transportation plays a major role in all countries regardless of their economical stature. It is a fundamental factor in ensuring businesses, which contribute to the wealth of a nation and acts as a stimulus for the economy can operate efficiently and continue thriving [1]. Hence, akin to communication, the efficiency and availability of transportation can be considered as a benchmark for the ability for nations to develop its economy. The reason for this is workers and employees are the backbone of large corporations and businesses of various size alike. Therefore, ensuring mobility is of utmost importance since efficient transportation can cut down time wasted on daily commutes and reduce physical and mental fatigue allowing the average worker or employee to be more productive at work while also allowing for more business meetings to be conducted. Various initiatives to achieve this have already been taken by governments such as laying roads and highway across nations, building various types of rail based transportation as well as airports which connect various cities and states. Most of these initiatives are usually concentrated in metropolitan areas where most businesses operate. However, with billions in funds being poured in to improve the facilities in these areas, various inadequacies persist such as traffic jams and bus or train delays. Furthermore, the congestion that stem from this issue such as sound and air pollution also pose as a threat to live ability of cities which in turn may affect the health of workers and their productivity, cause businesses to close for several days due to severe air pollution or cause people to shun the city altogether [2].

Therefore, a transportation solution that is efficient and makes use of existing infrastructure is proposed to minimize the economically harmful effects mentioned. The solution makes use of roads more efficiently by matching people that have a similar commuting pattern to carpool and sharing the cost between them without the hassle of finding people to carpool and conundrum of splitting travel costs. Finding a reliable yet comfortable mode of transportation especially in a city setting is somewhat difficult due to the various compromises that need to be made based on the transport mode chosen. For instance, driving and taking taxis or using ride-hailing applications may be comfortable modes of transportation within a city since the commuter has the flexibility to leave at any time while avoiding huge crowds [3]. However, a compromise needs to be made in terms of higher costs after factoring in maintenance, fare and parking costs besides getting stuck in unpredictable traffic. Meanwhile, though using public transportation such as trains may reduce the downsides of driving or taxis such as, they too have compromises to be made such large crowds and uncertain arrival times making it inflexible for the commuter to leave at a desired time. Furthermore, most users of public transportation would also have to do some walking making them subject to the elements such as rain since public transportations have set stops which may not be close to the commuter's destination.

Moreover, most road congestions are caused due to inefficient road usage instead of just the sheer number of people using roads to commute. For instance, most cars were found to be only occupied by a single occupant during road congestion [4]. Hence, road spaces are wasted since a single person occupies the amount of space that is available to transport at least 4 people. This practice in turn leads to unnecessary delays and pollution in cities. Recently, a number of ride-hailing services have started to offer taxi-like services where passengers wanting a ride would request for drivers through a mobile application. Though these services help reduce resources wasted such as time and cost of taxi drivers waiting to get flagged down by potential passengers, they still promote the same wastages associated with cabs which underutilize the maximum usage of the vehicle used for transporting and the road space filled by these transportations.

Other issues associated with achieving carpooling include a lack of security and flexibility of travel for the drivers involved in a carpool journey. The issue of lacking security is also prevalent in other private transportation forms such as ride hailing and cab services. Meanwhile, the flexibility of drivers when carpooling is compromised making most drivers unwilling to share their car even if it may reduce travel costs [5]. Therefore, the system will need to address these issues too to ensure that the service concedes to better adoption by the community targeted users of the system.

According to the problems discussed above, a ride-sharing application which allows drivers and commuters to share the cost of a journey should be able to provide the desired convenience and costs for moving about in a city. In contrast to ride-hailing applications such as Uber and Lyft which acts just like call taxis whereby a drive drives from any location nearby to fetch passengers to make profit, the application should use machine learning to learn daily user travel patterns and recommend "ride buddies" for which the ride is to be shared with. This also helps minimize pollution since drivers only pick up passengers en route to their destination while sharing the ride with as many people as possible. Moreover, the application should implement a messaging service to allow users to contact one another to discuss journey details such as ride costs, pickup points, alternative routes to be used and alert about delays.

This could also help on country tourism. The main pushing factors for a country tourism to grow is transportation and phone applications. Work by [6, 7] help companies travelling in the city with a cheaper cost and efficiency. This study proposes a web-service to connect the application and a machine learning module to provide suggested journeys based on users' travel patterns called the ORSUM.

The remainder of this paper proceeds as follows. Section 2 reviews system similar to the ORSUM. Section 3 presents the architecture and Section 4 presents the implementation of ORSUM. Section 5 discusses evaluation of ORSUM and finally Section 6 concludes with some indication for future work.

2. Related Areas

In designing the carpooling application, numerous carpooling and ride sharing applications are available to be used for free by the public. However, three applications stand out, namely, Grab, Lyft and Uber. Though the aims of these services are similar in which passengers are picked up and sent to their destinations like taxis, there exists several key differences in the approach to achieving this in terms of strategy, cost and services offered (Staff, 2017).

2.1. Grab

Grab started off as MyTeksi in 2011 in Kuala Lumpur and was later branded as GrabTaxi outside Malaysia. The mission of Grab according to Anthony Tan the CEO of Grab and Grab's co-founder, Tan Hooi Ling, was to improve the safety and accessibility of transportation, along with improving the lives of passengers and drivers [8]. In essence, Grab was aimed at efficiently matching passengers to taxi drivers to passengers. This was to ensure a steady stream of income for taxi drivers and shorter wait time for passengers wanting to take cabs. Furthermore, Grab was also able to establish a safer form of taxi since all drivers were to register with the company before being allowed to take passengers. Aside from Malaysia, Grab has also expanded into most other countries in the Southeast Asian countries such as the Philippines, Singapore, Indonesia, Vietnam and Thailand. Besides taxi-hailing service, Grab later expanded to include other services such as its ride-hailing service called GrabCar a variation of GrabCar that promotes carpooling called GrabShare and a car pooling service called GrabHitch which was introduced in late 2015. GrabHitch is the closes competitor to ORSUM due to sharing various features such as drivers being paid for variable costs such as fuel, car depreciation costs and distance travelled. Passengers are also able to view costs of the journey beforehand and book a ride between seven days and 15 minutes in advanced.

2.2. Lyft

Lyft was founded in 2007 as Zimride by CEO Logan D. Green and co-founder John Zimmer and is based in San Francisco, California. Similar to Grab, Lyft is also a smartphone-based ride-sharing and peer-to-peer transportation service that matches passenger who request rides to drivers [9]. However, instead of focusing on increasing income for taxi drivers, Lyft aims to make it easier for communities in corporate and educational institutions such as colleges and universities to commute. In August 2014, Lyft launched its own carpooling service called Lift Line. The idea began with the acquisition of Rover, a company building a transit application that matched users with the best route to their destination through shared transportation. Based on Lyft's observation, most people are willing to ride with strangers as long as it is cheap. However, previous efforts at car-pooling services such as Lyft Carpool which allowed regular drivers to earn money for providing a lift along their morning commute failed due to a lack of driver interest with Lyft citing "the time is not right now".

Lyft currently aims to make Lyft Line more efficient in matching and cheaper compared to other modes of transportation. Based on the research performed, several key factors influencing the domain of carpooling and transportation choices in general were identified. These include Affordability, Convenience, Duration of Commute and Security of each mode of transportation. Affordability was found to largely influence the selection of driving as a primary transportation mode due to fluctuations in global petrol prices and differing government incentives offered by various countries such as dedicated HOV lanes and tax rebates for purchasing energy efficient vehicles (EEVs). Besides that, several social concerns were also echoed in the literature researched to includes factors such as safety and to a smaller extent, social interaction with others during the journey. An analysis on similar systems had found that majority of the systems faced issues in terms of user adoption, especially amongst people who drive.

One of the main reasons cited for this issue was the lack of incentives for the drivers to want to carpool. However, it is important to note that though the companies that introduced their carpooling services were well-established and renowned in the ride-sharing industry, the carpooling services introduced were relatively recent being within the span of 1 to 3 years ago. Furthermore, another notable concern mentioned was security of the services. To address this, several measures had been put in place such as panic buttons and driver background verification. The latter would not be suitable for ORSUM due to the bureaucratic procedures that need to be faced by drivers to begin carpooling.

2.3. Uber

Much like Lyft, Uber too began in San Francisco, California a service allowing users to hail-rides through a smartphone application [10]. Similar to Grab and Lyft, Uber too began as a ride-hailing service to connect passengers and drivers while providing more details such as estimated time of arrival (ETA) to remove the question of when a ride will arrive (Pullen, 2014). Uber sources its revenue from processing ride payments by charging the passengers credit card while taking a percentage of 5%-20% before depositing the balance into the driver's account. In August 2014, Uber introduced a new service called UberPOOL. As its name suggests, UberPOOL is a carpooling service that matches passengers requesting for ride with drivers also heading in the same direction. The benefit of UberPOOL compared to the regular service is lower ride costs for passengers. This comes at the expense of slower matching of drivers and passengers, fixed route once the passenger is picked up and passengers requiring to be at the pickup point before the driver arrives.

3. Proposed ORSUM

ORSUM is similar to Uber and Lyft in the sense that it is categorized as a ride-hailing application that respond to requests similar to taxi calls. However, ORSUM is also different from other ride-hailing applications because it is actually a ride-sharing application that allows drivers and commuters to share the cost of a journey through shared travelling. The proposed ORSUM application is designed with four distinct modules; (1) ORSUM Machine Learning module (ORSUMML), (2) Azure Database, (3) ORSUM Web Application, and (4) ORSUM Application. The machine learning module runs as a standalone Django web application that is separate from the ORSUM Web Application and only interacts with the ORSUM Web Application.

3.1. ORSUM machine learning module

The Orsum machine learning module is responsible for providing suggested trips to users based on their respective location history. As mentioned earlier, the module uses a mean-shift clustering algorithm to cluster users with a similar journey trend. Based on the results of the clustering algorithm, the module returns a list of user ids that have a similar travel pattern to the user requesting a suggested trip. The reason for separating the Orsum machine learning module from the Orsum web application is firstly due to performance. This is due to the clustering algorithm having to consume a significant amount of computing resources. Hence, modularizing the component would help to offload resources from the Orsum Web Application that will be handling requests more frequently.

The Orsum Machine Learning module was also separated due to compatibility and ease of integration and maintenance. Due to high complexity of the clustering algorithm and machine learning algorithms in general, high-level languages such as Python are favoured by the community and therefore have dedicated libraries that help

build the models. Hence, the module was written in Python usin Django as a separate web service due to difficulty and issues integrating with ASP .NET Core applications. Furthermore, separating the two applications also allow for improvements to be made to the machine learning algorithm without having to make amendments to the Orsum Web Application allowing for easier testing and deployment.

Among the core issues to be resolved is the type of machine learning algorithm to be implemented within the business logic layer of the system. Machine learning is the use of statistical models that receive data as inputs and perform calculations based on the inputs to output useful information. In the case of ORSUM, an unsupervised machine learning algorithm is to be applied. As its name suggests, unsupervised machine learning algorithms do not require human intervention to determine how the outputs of the model are to be classified. Instead, the algorithm itself is able to perform classification and requires the user to derive information based on the outputs. From this, it is clear that the data input required to build the model needs to first be identified. Next, how the output from the algorithm is going to be analysed and used for the system will also need to be figured out. Factors such as these may influence the effectiveness and correctness of the algorithm in the system's processes.

Furthermore, the algorithm selected for the machine learning module should not only be judged by correctness, but also by its performance. This is to guarantee that the application can get rides faster compared to other forms of transportation that are cheaper alternatives. To ensure better performance, it would be preferable to try performing the calculations for the algorithm with the use of GPUs. The algorithm used for ORSUM's machine learning module is the Mean-Shift clustering algorithm. Mean-Shift is an unsupervised hierarchical clustering machine learning algorithm. A clustering algorithm is used to divide up a dataset into several circles of clusters called kernels. All machine learning algorithms require for the models to be built by a process called training where a large amount of data is containing inputs and the desired outputs are passed into the algorithm allowing it to build a model.

As an unsupervised algorithm, the outputs for each input for algorithm does not need to be set by beforehand whilst training the model [11]. This means that the algorithm is able to derive the output by itself without specifying the correctness of the output. Clustering is often confused with classification. A key difference between the two types of machine learning algorithms is that classification attempts to classify data fed into known or predefined classes that the model has already been trained with. Clustering meanwhile, attempts to link the data with similar features as a cluster while maximizing the differences of data with dissimilar features to discover relationships within the data fed.

Furthermore, being a hierarchical clustering algorithm allows for the number of clusters to be deduced by the algorithm without having to specify the number of clusters to generate from the dataset. As shown by the diagram below, the algorithm was able to deduce three clusters from the dataset input by itself without specifying th number of clusters to create. Fig. 1 shows the sample mean-shift algorithm clusters from [11].

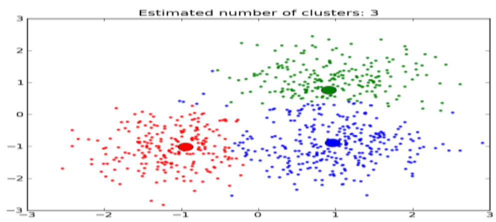

Fig. 1: Sample mean-shift algorithm clusters [11]

As its name suggests, Mean-Shift forms clusters by adjusting th centroid based on the mean of each cluster whenever new data I added to the model. It begins by accepting each data point as a cluster of its own. Hence, at the beginning of the algorithm, all data points begin as an independent cluster. However, each cluster in Mean-Shift has a bandwidth which represents a diameter surrounding each data point of the dataset and is set beforehand. The bandwidths are used for clustering data points with similar features into a single cluster by converging the points. This is achieved by having data points with overlapping bandwidths grouped as a single cluster. Each cluster has a centroid which is calculated based on the mean of all data points within a cluster. Every time a new data point is added to a cluster, the centroid for said cluster is recalculated accordingly. This process is performed recursively until there is minimal to no shifts in clusters positions or cluster centroids.

As shown in Diagram 2, the red points represent the centroid of clusters. At the beginning, all data points are considered as separate and independent clusters. After the first iteration of the algorithm, several data points with overlapping bandwidths are then pulled together forming a single cluster with fewer number of centroids. After every iteration, the number of centroids and clusters are considerably reduced. After 5 iterations however, only three centroids of three cluster remain.

The way the algorithm is to be used for ORSUM is by using various transport related data as inputs for the algorithm. This includes data such as average time of departure, destination of commute and origin of commute being used as arguments for Mean-Shift. The result of doing this is a 3-D model of cluster is shown in Fig. 2.

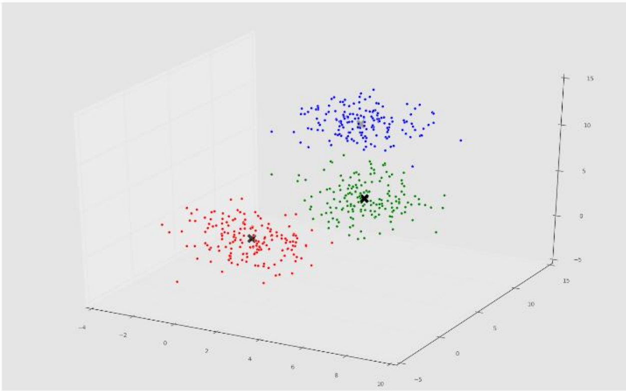

Fig. 2: Sample 3-D hierarchical modelling

Based on the arguments supplied into the algorithm, several clusters containing people of similar commute patterns are to be formed. Whenever a new request is made or the average departure time for a person within a given cluster is getting close, the system will proactively try to match users within the same cluster to enable shorter wait time to ensure users with most similar commuting pattern is matched for a journey.

The reason for choosing the Mean-Shift over the K-Means algorithm is due to the need for a hierarchical clustering for the matching similar users of the system. With K-Means, K number of clusters will need to be defined for the algorithm beforehand. This would lead to inaccuracies since the number of clusters that need to be formed cannot be predetermined accurately due to fluctuating features of the users whenever new users are added to the system. However, using Mean-Shift will allow for the number of clusters to be formed automatically. The only argument that needs to be supplied is bandwidth of cluster centroids which can be optimized based on another algorithm used to calculate the optimum bandwidth size relative to the number of users registered to the system. This allows more flexibility for the formation of clusters as required by the system for accurately clustering users. In simpler terms, bottom up approach such as the Mean-Shift algorithm is required to try matching similar users of unknown number of categories as opposed to the top-down approach of the K-Means algorithm.

A drawback from using Mean-Shift compared to K-Means is the algorithm performance being more resource intensive than K-Means which affects its scalability as data grows. However, this will need to be dealt with through system software and hardware optimizations since algorithm accuracy is more important in the case of ORSUM in achieving the system’s objectives

3.2. Azure database

The Azure Database holds all records and data of the system such as location and journey history, trips and user information. The Azure database never interfaces directly with the ORSUM Application, instead only through the ORSUM Web Application and the ORSUM Machine Learning module. All queries are executed through the use of stored procedures stored within the Azure Database and which are invoked by the two web applications it interfaces with.

3.3. ORSUM Web application

The Orsum web application is an ASP .NET Core web application that is designed for cross-platform usage. It serves as the backe connecting Orsum application to the Azure Database. As such, all queries from the Orsum Application are routed through the web application before being return to the user. Aside from that, the web application also performs validation to ensure data integrity to avoid issues such as duplicate or invalid data. Furthermore, the web application is also responsible for handling device registrations to the system as well as performing sending of push notifications when users send or receive messages. This is handled natively by Orsum Web Application due to limitations of Azure when handling cross-platform notifications. This is done by monitoring the devices registered to each user and removing those with duplicate devices. Moreover, the web application also coordinates the creation and scheduling of journeys, trips and suggested trips. Furthermore, a web service preferably one that can cater to multiple mobile platform needs to be identified and implemented. Moreover, a mobile platform for the mobile application needs to be identified. The platform selected should be one where it would be easy to port over to other mobile platforms while not compromising on performance of the application to ensure good usabil-

ity. As such, C# will be used as the primary language due to various tools available to port the program to other mobile platforms. The application will be developed natively instead of cross-platform to ensure optimized performance when the application is run.
Moreover, the system will be required to not only capture data about a user's daily movements but also use the data to recommend rides that might be useful for the user's commuting needs. Meanwhile, the same tracking is to be done for users that drive with the difference being drivers get notified when passengers request for a ride to be shared and is able to accept or decline the request. The system needs to be optimized well enough such that the entire process from location capture, accepting requests, processing optimum ride participants and suggestion of ride participants are done within an acceptable timeframe to ensure usability of the system.

3.4. ORSUM application

The Orsum application is a cross-platform application that runs on Android and Windows UWP. It is built using Xamarin Forms a serves as the interface between users and the system. All functions of the system are initiated through Orsum Application such as crating journeys, planning trips, receiving suggested trips, location updates, messaging and finding rides. The main goal of the Orsum Application is to fetch data from the Azure Database through the Orsum Web Application or Orsum Machine Learning module before displaying to the user and to capture data such as journey, location and user personal information.
Though the system is usable by virtually all demographics, it is targeted towards people living in densely populated areas such activities. To be more specific, the system targets those whom frequently commute during rush hour in cities when transportation u sage is at its peak. These mostly include office-working adults that drive to work within the lower end of the age spectrum such as those in their 20s and early 30s. They are presumed to mostly travel alone due to not having started a family and would therefore also be more willing to share their vehicle with others travelling in the same direction.
Furthermore, the same would also apply for potential passengers of the system since these demographics of people are usually just beginning their professional lives and are looking for various means to save on travel and living costs. In addition, some of those belonging to this demographic may not be able to afford a car just yet and are financially obligated to use other modes of transportation that are less convenient than driving.
Moreover, due to presumably not yet starting a family, these group of target users would probably have fewer commitments allowing them to be more attuned to the idea of carpooling. To develop carpooling system that uses machine learning to pair passengers to drivers based on prediction of routes being taken and recommends a split for the cost of the journey after factoring in fuel, luggage load, number of passengers and toll costs between the two.

4. Prototype Implementation

The Sign-up page is the first page used by new users to create their account. For account creation, users are required to input their personal details such as name, email, phone, password and state that they reside in. Each entry is checked for correctness upon completion. An input that does not satisfy the constraints of the entry will be colored in red to indicate an error. This is accomplished with the use of the FormValidator class. Furthermore, the sign-up button is not enabled until all field have been filled. If a field has errors and the sign-up button is clicked, an error message is displayed and the entry field highlighted. If there is already an account associated with the email entered, a message is displayed requesting the users to login as an account already exists and the Login page displayed. The Login page is the first page to load for users that have not been signed in. From here, users are required to enter their credential (email and password) to log into the system. If incorrect credentials are entered, the system displays an error message requesting the user to re-enter credentials. If the email entered is not in the correct format, an error message is displayed notifying the user. The login button is only displayed once both entries have been filled.
The Profile page displays user personal details that were captured during registration. From here, users are able to review their personal information. At the bottom of the page is a button that allows users to change their password. Tapping the button will display the change password page. Fig. 3 shows the interface for the Profile page.

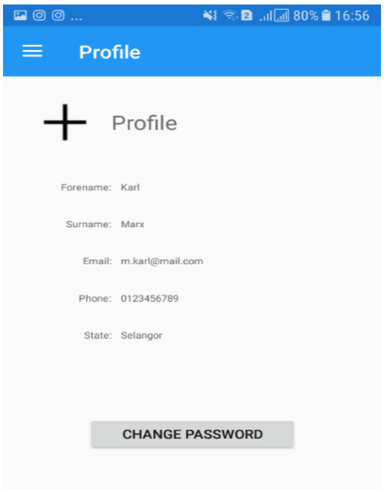

Fig. 3: Profile

The Today page is the first page users are greeted with upon logging into the system. The today page displays a list of upcoming trips for the current day. Aside from that, the today page also displays a suggested journey for the current day for the current day. The suggested journey is populated by leveraging the functionality of the ORSUM machine learning module. This is done by obtaining a list of users with similar travel pattern using the historical location data of the users. Users can join a suggested journey by tapping on the suggestion area and view more details of an upcoming trip by tapping on it. Fig. 4 shows the interface for the Today and New Ride page.

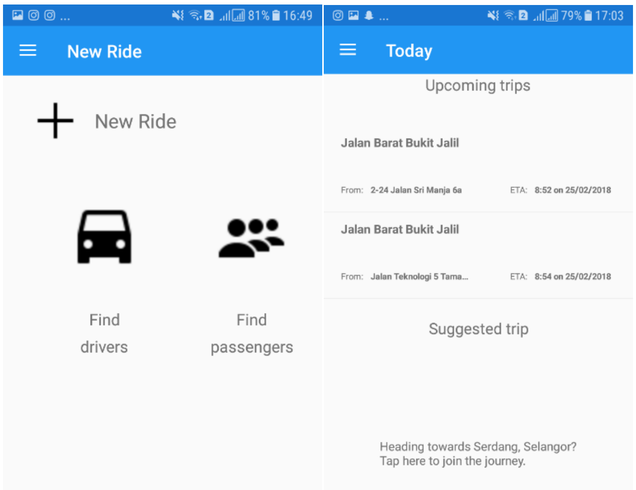

Fig. 4: Interface for Today and New Ride

The New Ride page is the first page involved in the manual creation of a journey. On this page, users select whether they would like to find for passengers as a driver or are looking for drivers if they are a passenger. Selecting the Find for Passengers option would allow users to input the number of passengers they are expecting to carry on the Create Journey page. In this page, users specify the location information of the journey to be created. This includes the starting point and destination of the journey. As users complete either fields, a map shown on the bottom half of the page drops a pin on the location specified based on the results received from Google's Geocoder API service.

Users may also specify their current location as the starting point by tapping the current location icon beside the starting point entry. Once both entries have been filled, the Next button is enabled allowing users to proceed with the following step in journey creation. If no route is found between the starting point and destination specified, an error message is displayed. The same goes if a location entered in either fields cannot be found. The next page allows users to input timely information about the journey being created. This includes the date of the journey and the time the user expects to reach the destination specified. If the Find Passengers option was selected from the New Ride page, an option to input number of expected passengers is also made available to the user. Fig. 5 shows the interface for creating the journey and information on existing journey.

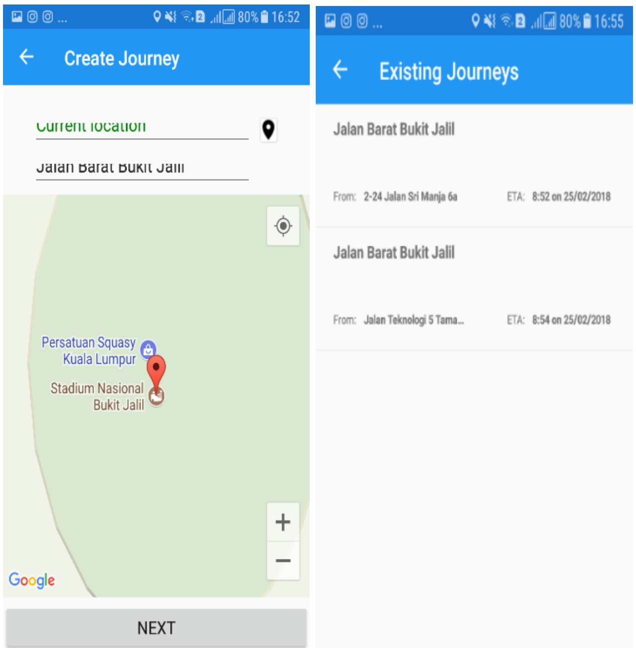

Fig. 5: Create Journey Page

The Journey page displays information about a journey. The left section of the page displays a map with an overlay of the route for the journey. The right section displays detailed information about the journey such as the starting point, ETA, total passengers, if the user is a driver for the journey and journey status. Below that, is a list of names of other participants of the journey along with their respective starting points and destination. On the bottom section of the page, a chat button is available which opens the chat interface. If the user is the driver of the journey, a start destination button is also available at the bottom of the page which begins navigation to the starting points of other participants of the journey by opening setting directions on the native maps applications. The Existing Journey page displays a list of journeys related to the user. Among the information displayed for each trip is the destination, starting point and estimated time of arrival (ETA) for the journey. For journeys with the status "Pending", the background of the journey is highlighted in blue. Meanwhile, passengers of a journey that has already been completed are highlighted in green. This indicates that the journey is awaiting a rating. Fig. 6 shows the interface for the Journey and Existing Journey page.

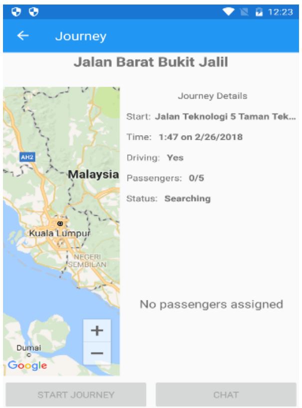

Fig. 6: Journey page

Next, the Request page is similar to the Existing Journey page except that it is used to display a list of trips where drivers are searching for passengers to join the trip. Tapping on a request display more details about the trip. If the user is interested in joining the trip, they may request to join the trip. The rating of the driver is displayed on the top right corner of each request in the list. Besides that, all other information displayed is similar to the Existing Journey page for the trips displayed such as destination, starting point and estimated time of arrival (ETA). Fig. 7 shows the interface for the Request page.

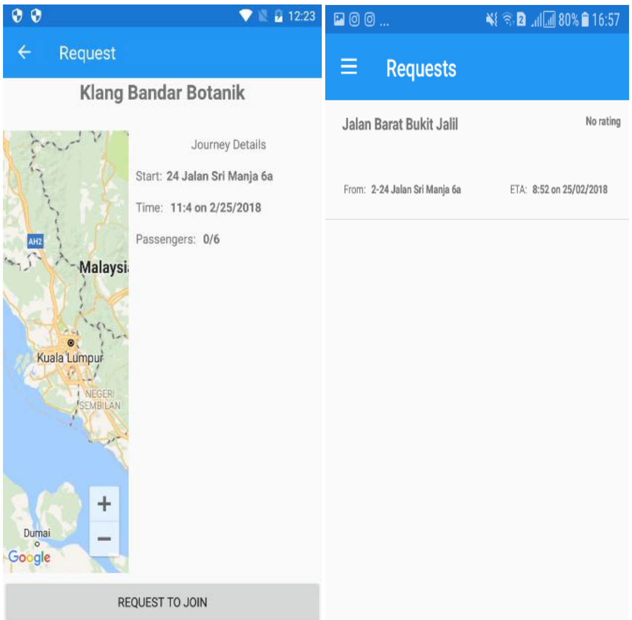

Fig. 7: Request page

The request page is displayed when a passenger selects a request from the requests page. Similar to the journey page, the left portion of the page displays a map with an overlay of the route to be taken by the trip. The right portion meanwhile includes information about the trip such as starting point, expected time of arrival (ETA) and total passengers of the trip while the top of the page displays the destination of the trip. At the bottom of the page is a "Request to join" button that allows passengers to send a request to the driver of the journey about the intent to join as a passenger.

5. Evaluation

ORSUM was evaluated qualitatively using questionnaire. The first section will be used to describe the demographics of the questionnaire's respondents. Next, respondent sentiment towards carpooling and the motivations and disincentives for carpooling between those who drive and take public transportation/ use other modes of transportation will be discussed. An indication of carpooling satisfaction will also be derived from the questionnaire. Furthermore, the data will also be used to understand how drivers, car poolers and those whom rely on public transportation perceive their modes of transportation with regards to the main influencing factors identified earlier. Besides that, a discussion of how respondents perceive all key transportation modes will be summarized. Lastly, the importance of each of the major influencing factors identified when choosing a transportation mode will be discussed.

All in all, 84 responses were collected for the questionnaire. Most of the respondents were male belonged to the age bracket of 18-29 years old. Furthermore, most of the respondents were students followed by those working. The rest were followed by a small minority of those unemployed, housewives and retirees. As targeted by the questionnaire, most of the respondents were those dwelled in urban areas with followed by a small margin who stay in suburban areas. Most of the respondents had also recently carpooled with a quarter of the total respondents having carpooled within the last 3 months.

This indicates a decent number of respondents comfortable with carpooling. Fig. 8 shows respondents' feedback on the experience

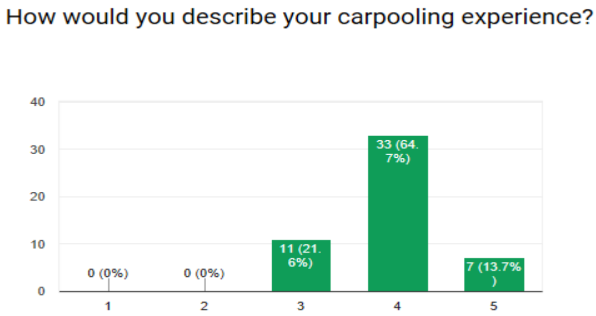

Fig. 8: Carpooling experience

Based on the data gathered, most respondents were found to be in favor of carpooling followed by a smaller percentage that would consider carpooling and minute portion saying no to carpools. For those whom agreed to carpool, the motivations can be summarized into three broad categories: Environmental, Cost and Time. Among the three, Cost was cited by most respondents as the reason for favoring future carpools in order to save fuel costs when transportation is shared with others. However, the difference between the three factors were minimal indicating the three factors being equally important for those willing to carpool. Environmental factor was also mentioned second most frequently among the three factors with majority participants citing a personal responsibility to reduce sound and air pollution. Lastly, many felt carpooling saved them time possibly due to a lack of parking facilities in cities.

Meanwhile those who considered future carpools said so primarily due to a matter of convenience such as a difference in time and destination. Fig. 9 shows respondents' satisfactory level on carpooling.

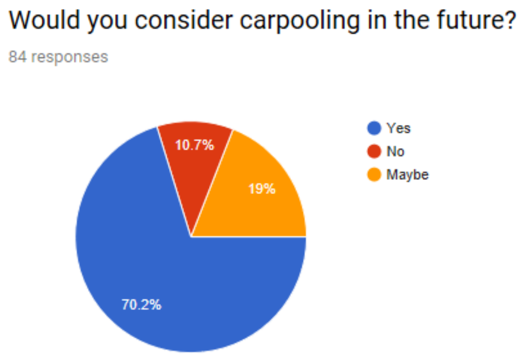

Fig. 9: Carpooling experience

The reason for refusing future carpools were only due to difficulty finding others to carpool. However, only one response echoed concerns about personal safety whilst carpooling. On a scale of 1 to 5, those whom had carpooled within the last three months were generally positive about their experience carpooling. Most of them were found to be somewhat satisfied with the experience followed by slightly more than a quarter whom had a neutral stance towards their carpooling experience. A small number of respondents were also very pleased with the experience with none rating the experience negatively with a rating of 1 or 2.

In evaluating the perception of drivers, respondents are chosen among public transportation users such as those using bus and train and those whom carpool on their own modes of transportation. Most of the questionnaire respondents were of those whom primarily drive making up 72.6% of total respondents. This is followed by public transportation users with Train and Bus users comprising of 12.9% and 9.7% respectively. The number of respondents whom carpool or cycle/walk meanwhile only constituted 3.2% and 1.6% respectively.

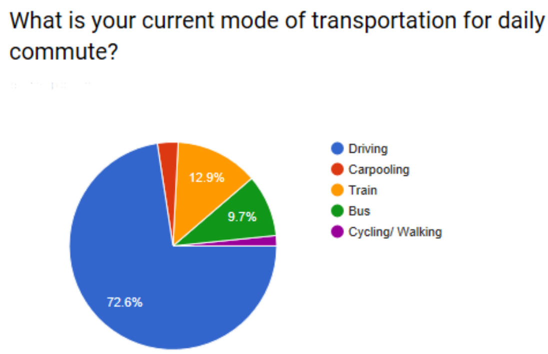

Fig. 10: Respondents' daily mode of transportation

All in all, the system passed the unit tests that had been conducted. However, several issues were faced during the User Acceptance Test (UAT). Among the issues highlighted were the displaying of maps on the journey and request pages. This was due to feedback received whereby pins were not informative in giving an overview of the trips route. To overcome this, an overlay was used on the map to highlight the route to be taken by the driver for a trip with the help of the Google Directions API service. Another UAT feedback was the back navigation of pages when carrying certain tasks such as when creating a new journey. This led to confusion as the context of the task was lost when the user was trying to navigate back to some of the previous pages. To overcome this, a new algorithm was used to manage the navigation stack when certain tasks were carried out to limit the pages users are able to navigate back to.

Moreover, some of the error messages have been revised to be made more concise and user-friendly. Cosmetic changes were also made to the journey and request pages to make the size of the map displayed smaller in relation to the journey details pane. However, an issue that could not be resolved was the performance of the suggested trip feature due to Azure hosting plan limitations and limitations to the algorithms accuracy and scalability due to a lack of data.

6. Conclusion and Future Plan

This paper has presented a ride-hailing application called ORSUM, which capitalizes on machine learning approach to learn users' daily travel patterns and recommend "ride buddies" for which the ride is to be shared with. Among the things achieved at the end of the project include the development of a ride-sharing system used to facilitate finding or rides and passengers for carpooling. Ths included the development of a cross-platform application that serves as the interface between users and the ORSUM system, a web-service to connect the application and a machine learning module to provide suggested journeys based on user travel patterns. With regards to the investigation carried out for the project, most aspects of what was to be achieved was able to be collected successfully. However, an area with inadequate research is the motivation behind potential users especially drivers wanting to carpool, more specifically the economical motivations for doing so. Hence, further research needs to be carried out to investigate a fair method of calculating fares for journeys to ensure users are drawn towards using the system.

Aside from that, an area that needs to be improved upon is the machine learning module and the design of the model itself. This includes changes to be made to the algorithm itself to improve the scalability of the system as well as parameters fed into the model to ensure the outputs of the algorithm are more accurate. Furthermore, changes need to be made to the Orsum front-end application to make it more intuitive and attractive to the end-user. Moreover, thought the accuracy of the algorithm was tested to be accurate on a smaller scale of data. It is unable to be completely demonstrated on a larger scale due to a lack of data for testing the module.

Acknowledgement

This work is supported by Asia Pacific University of Technology and Innovation, Malaysia.

References

[1] B. Herrendorf, J. A. Schmitz, Jr, and A. Teixeira, "The role of transporation in us economic development: 1840–1860," International Economic Review, vol. 53, no. 3, pp. 693–716, 2012.
[2] S. K. Guttikunda and R. Goel, "Health impacts of particulate pollution in a megacity—delhi, india," Environmental Development, vol. 6, pp.8–20, 2013.
[3] L. Redman, M. Friman, T. G¨ rling, and T. Hartig, "Quality attributes of public transport that attract car users: A research review," Transport Policy, vol. 25, pp. 119–127, 2013.
[4] U. T. T. Force, "The high cost of congestion in canadian cities," Council of Ministers Transportation and Highway Safety, 2012.
[5] T. Aprikyan, "New ride-share apps offer flexible commutes, but at what cost?"https://q13fox.com/2017/04/21/ new-ride-share-apps-offer-flexible-commutes-but-at-what-cost/.
[6] V. Kasinathan, A. Mustapha, Y. C. Seong, and A. Z. Z. Abidin, "Foot-print: Tourism information search based on mixed reality," 2015.
[7] V. Kasinathan, A. Mustapha, and T. Subramaniam, "Smartg: Sponta neous malaysian augmented reality tourist guide," in IOP Conference Series: Materials Science and Engineering, vol. 160, no. 1. IOP Publishing, 2016, p. 012103.
[8] M. LIN and C. DULA, "Grab: Discovering new frontiers for growth in the southeast asian sharing economy," 2016.
[9] A. Henao, Impacts of Ridesourcing-Lyft and Uber-on Transportation Including VMT, Mode Replacement, Parking, and Travel Behavior. University of Colorado at Denver, 2017.
[10] A. Hartmans and N. McAlone, "The story of how travis kalanick built uber into the most feared and valuable startup in the world," Business Insider, 2016
[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al.,"Scikit-learn: Machine learning in python," Journal of machine learning research, vol. 12, no. Oct, pp. 2825–2830, 2011.